通过训练可微绑定函数实现绑定反演

Mathieu Marquis Bolduc
mmarquisbolduc@ea.com SEED
Electronic Arts 奥斯汀，得克萨斯州，美利坚合众国

Hau Nghiep Phan
hphan@ea.com SEED
Electronic Arts 蒙特利尔，魁北克，加拿大

arXiv:2301.09567v1 [cs.GR] 11 Jan 2023

摘要

绑定反演旨在创建一种方法，以找到最能近似给定输入网格的绑定参数向量。本文提出通过训练多层感知机来近似绑定函数，从而获得可微绑定函数，以此解决这一问题。随后，可以利用该可微绑定函数训练绑定反演的深度学习模型。

CCS 概念

• **计算方法** → **机器学习**；**动画**。

关键词

绑定反演、神经网络、计算机动画 n

ACM 引用格式：

Mathieu Marquis Bolduc 和 Hau Nghiep Phan。2022年。通过训练可微绑定函数进行绑定反演。载于*SIGGRAPH Asia 2022*技术交流（*SA '22* 技术通讯），*2022*年*12*月*6*日至*9*日，大邱，大韩民国。ACM，纽约，纽约州，美国，4页。https://doi. org/10.1145/3550340.3564218

1 引言

作为传统动作捕捉（追踪关节位置）的替代方案，可以直接从各种捕捉系统（例如[Fyffe et al. 2015；Ghosh et al. 2011]）获取三维和四维网格数据。此类系统的一个设计动机，是以高保真度捕获软体变形。理想情况下，这些三维数据应当经过绑定处理，即获得与之最佳对应的一组动画参数（例如动画工具图形用户界面参数）。这样，三维动画师便可以使用自己的工具编辑姿势，也可以在针对特定绑定开发的各种工具和软件中使用捕获数据。如果绑定逻辑是一个以动画参数 [Hahn et al. 2012], 或绑定参数为输入、输出三维网格的函数，那么，求得与某个三维网格对应的绑定参数，就是对该函数进行反演的问题。需要注意的是，我们希望反演的绑定逻辑可能既不是单射，也不是满射：相同的网格可能由不同的绑定参数组生成；而且，绑定函数几乎不可能完美输出任意给定的可捕获网格。因此，在实践中，不应将绑定函数视为双射，也就是说，它无法被完美反演。因此，绑定反演问题就转化为寻找一个函数，使其输出绑定参数；将这些参数输入绑定逻辑后，能够使输入网格与绑定后的网格之间的差异最小。

许可免费制作本作品的全部或部分内容的数字或纸质副本，供个人或课堂使用，但不得为营利或商业目的制作或分发副本，且副本首页须附上本声明和完整引文。对于本作品中由作者以外的其他方拥有版权的组件，必须尊重其版权。注明出处后允许摘录。若要以其他方式复制、再版、发布到服务器或分发给列表中的用户，须事先获得具体许可和/或支付费用。请向 permissions@acm.org 申请许可。*SA '22* 技术通讯，*2022*年*12*月*6*日至*9*日，大邱，大韩民国© 2022年版权归所有者/作者所有。出版权已授权给 ACM。ACM ISBN 978-1-4503-9465-9/22/12...$15.00
https://doi.org/10.1145/3550340.3564218

2 先前工作

近年来，绑定反演一直是一个备受关注的课题。鉴于基于雅可比矩阵的方法的性能开销会随着训练样本数量的平方增长，[Holden 等人 2015]提出了一种更节省内存的方法：使用动画师提供的数据训练多层感知机，以最小化给定输入网格时输出绑定参数之间的差异。 [Holden 等人 2017] 还提出了一种基于高斯过程回归的方法。 [Racković 等人 2021] 将顶点与相应的绑定控制参数聚类，使得每个簇都可以通过高斯过程回归进行反演。 [Rackovic 等人2021] 通过使用二阶多项式对绑定建模，并使用列文伯格–马夸尔特优化对其求逆，改进了绑定的线性近似。 [Gustafson 等人 2020]以解析方式创建了一个针对推理优化的绑定版本，从而加快基于雅可比矩阵的绑定反演方法。

3 所提出的方法

在本研究中，我们假设绑定逻辑 $L(r)$（公式 1）是复杂且非线性的，它根据绑定参数值 r返回网格 x。我们使用的面部动画绑定正是如此，因此无法通过线性方法解决绑定反演问题。我们还假设绑定包含的参数过多，无法通过基于雅可比矩阵的方法在实际应用中求解。我们注意到，我们的绑定无法像 [Racković 等人2021] 提出的那样完全聚类，因为它们不符合该研究中的聚类假设。尝试使用先前发表的方法（例如 [Holden 等人2015], ）通过捕获会话中的网格反演绑定时，我们遇到了若干困难，下面将详细说明。

$$L(r) = x \tag{1}$$

$$L^{-1}(L(r)) = r \tag{2}$$

$$L(\hat{L}^{-1}(x)) \approx x \tag{3}$$

3.1 反演绑定函数时遇到的问题

3.1.1 非满射性。 来自绑定动画的有限数据并不总能很好地泛化到训练期间未观察到的新姿势。即使训练集内容全面，由于我们的绑定函数并非满射，仍有一些捕获网格不存在于绑定网格空间中，因此无法由基于绑定动画构建的训练数据集表示。

如第 3 节所述，在这些不存在精确解L^{-1} 的情况下（式 2），我们希望转而求得一个近似逆向绑定函数 $\hat{L}^{-1}(x)$，它在应用于绑定逻辑 L 时，能够提供最能近似输入网格 x 的绑定参数向量 r（式 3）

3.1.2 非单射性。我们的绑定函数不是单射的，也就是说，可能存在一组不同的输入绑定参数，却生成相同的网格。这意味着，使用绑定参数上的回归损失来训练深度学习模型，以反演此类绑定，会受到噪声干扰，因为训练数据集中，相同或非常相似的输入会对应不同的真实值。在实践中，该损失产生的梯度会使模型学会输出这些真实值的平均值，但无法保证这些平均值是良好的解。

3.1.3 绑定参数分布。在实践中，绑定参数向量 r 中的各个值并非同等重要。某些绑定参数对生成网格的影响比其他参数更显著。尽管我们希望学会预测所有参数，但若要尽可能减小网格 x 与 $L(\hat{L}^{-1}(x))$之间的感知差异，有些参数即使误差率较高也可以接受。然而，各参数的重要性难以衡量，且容易受到人为偏差的影响。

3.2 获取可微绑定函数

我们提出的方法基于这样一个想法：只需对训练过程做出同一项改变，就能解决上述三项难题。我们提出训练一个深度学习模型来反演绑定函数：将针对输出绑定参数向量 r的损失，替换为针对网格 $L(\hat{L}^{-1}(x))$ 的损失；该网格是将绑定参数向量 L^{-1} 输入绑定函数 L 后生成的。通过获得绑定函数 L 的可微近似 L_d，即可在实践中实现这一方法。随后，在训练过程中，可以采用编码器-解码器的方式（图 2），在反演后的绑定模型之后使用该近似来生成输出网格。然后，可针对该生成网格计算损失，并将其反向传播至反演后的绑定模型。

绑定函数通常被视为黑箱 [Holden et al. 2017；Seol 和 Lewis 2014]，可以对其求值，但无法进行解析操作。此外，绑定函数 L 可能无法在整个输入域 $\bar{r}$上推导出来。不过，我们可以训练出绑定函数的全可微近似 $\hat{L}_d$。我们提出使用一组随机选取的绑定参数向量训练一个解码器形状的多层感知机，这些向量已通过动画软件中的绑定逻辑求值（图 1）。与其逆函数 L^{-1}不同，绑定逻辑 L 被假定为真正的函数，因此，训练深度学习模型来近似它，相比近似其逆函数，可能会容易得多。近期研究进一步介绍了如何更好地训练绑定函数 [的近似模型 2020]Song 等人。

3.3 训练绑定反演模型

有了绑定的可微近似，我们便可仅使用网格数据来训练绑定反演模型，而无需对应的绑定参数向量真实值（图 2），从而解决本节开头提出的三个问题。为了获得

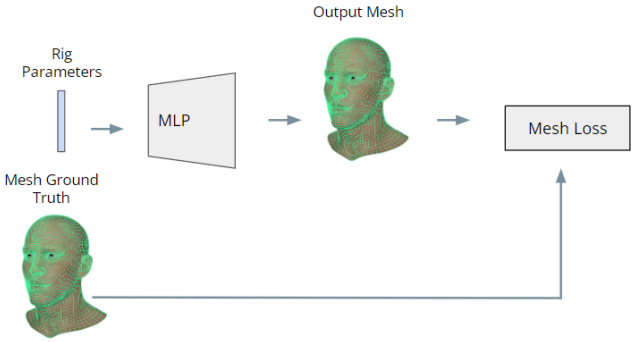

图 1：学习绑定函数的近似。

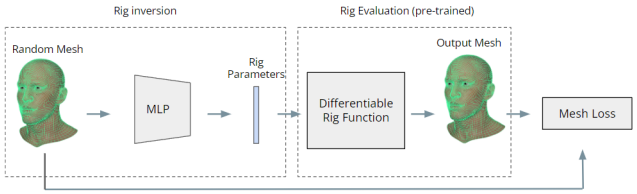

图 2：使用可微绑定函数近似来训练绑定反演模型。

良好的泛化能力，这些训练数据应进行随机增强（第 3.1.3 节），以尽可能多样地提供带噪声但仍可辨认的网格，并涵盖尽可能丰富的合理姿态。关于我们提出的最后一个问题（第 3.1.3 节），我们假设网格拓扑在感知上最关键的区域中，相对于顶点数量进行了过采样。这一假设带来的有利结果是，最关键绑定参数上的误差会在网格上造成更大的损失。

还有一些其他属性需要考虑。正如 [Holden 等人 2015] 所指出的，我们希望得到艺术家认为合理的绑定参数。因此，绑定参数的稀疏性是一个理想属性，因为艺术家希望尽可能减少需要操作的参数数量。在绑定反演模型中使用 ReLU 激活函数，是鼓励稀疏性的好方法 [Goodfellow 等人 2016]。最后，我们也希望动画后续帧中的绑定参数在时间上保持连贯；如果绑定函数不是双射，这一点就无法得到保证。为此，我们可能需要在绑定反函数网络中强制实现局部 Lipschitz 连续性。噪声增强是实现这一点的有效方法 [Usama 和 Chang 2018]。

4 实验结果

我们展示的所有定量实验结果均来自 EA 内容创作团队使用的一套面部动画绑定。该绑定有 137 个输入参数，可输出 628 个权重；这些权重通过矩阵乘法进行线性组合，生成一个含有 8447 个顶点的网格。此外，使用动画而非捕捉数据进行的定性测试证实，所提出的方法也适用于其他类型的动画绑定（图 4）。

4.1 可微绑定函数

我们从实验绑定中生成了 100000 组随机绑定参数值。如第 3 节所述，稀疏性是理想的，因此我们按照均匀分布，同时选择非零值的数量及其各自的数值。我们将这些随机绑定参数激活值输入 Maya 的 Python 脚本系统，以获取相应的权重，进而得到网格输出。该数据集的 10% 留作测试集，其余 90% 用作训练集。此前的研究，例如 [Holden 等人 2017]，指出这类训练数据集的规模必须以大得不切实际的方式增长。只有当所有输入和输出都彼此相关时，这种说法才成立。实际上，这种情况很少发生，本实验所用的绑定也并非如此，因此数据集的规模可以控制在实用范围内。由于该数据集是在均匀分布下随机生成的，训练期间不使用验证集。

ˆ用于对 L_d 建模的架构是一个多层感知机，包含两个隐藏层，每层有 1024 个参数，输出 628 个因子。隐藏层使用带泄漏的 ReLU 激活函数 [Nair 和 Hinton 2010]，而输出层使用 Sigmoid 激活，将输出限制在[0,1] 之间，即权重范围。我们不使用归一化，因为这会严重损害实验结果。模型使用 Adam 优化器进行训练，初始学习率为 1e-4；每当训练损失连续 20 个训练轮次处于平台期，学习率就减半，直到低于 1e-6。混合变形因子与矩阵相乘，损失函数采用生成网格顶点位置的均方误差损失。所有结果在不同的随机权重初始化下均保持一致。

由于我们用于实验的绑定不同寻常地并非黑箱，因此我们还将结果与使用 PyTorch™ 实现的该绑定函数的可微版本进行了比较。

表 1：绑定逻辑近似结果，显示以毫米为单位的顶点误差。

模型	平均顶点误差	最大顶点误差
程序化	**0.22**	**1.4**
多层感知机模型	0.78	5.9

表 1 比较了训练得到的绑定逻辑近似与程序化绑定逻辑近似的测试结果。学习得到的近似相当精确，平均误差不到一毫米。程序化方法产生的非零误差，可归因于 PyTorchTM 与 MayaTM 实现之间曲线求值的细微差异。

4.2 逆向绑定近似结果

训练后的可微绑定函数和程序化可微绑定函数均用于训练逆向绑定函数近似模型。两种情况下所采用的架构都是编码器形状的多层感知机，包含 4 个隐藏层，每层有 [2048, 1024, 512, 256] 个参数，并采用 Leaky ReLU 激活函数；输出层则采用 tanh 激活函数，以匹配绑定参数的取值范围。必须避免逆向骨架模型向绑定模型近似器输出超出流形范围的值，否则会导致不可预测的行为。与绑定函数模型一样，我们不使用归一化层。

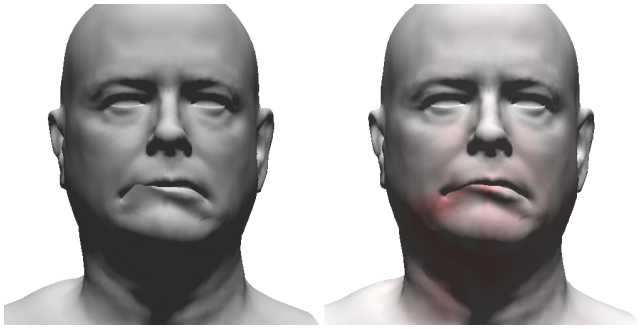

图 3：左：不考虑绑定逻辑而随机组合混合形状，会生成可辨认但带有噪声的网格，类似于采集过程或内容生成模型产生的不完美网格。右：使用网格损失训练逆向绑定逻辑，使其找到应用于绑定逻辑时最能近似输入网格的绑定参数，这一过程也兼具去噪作用。

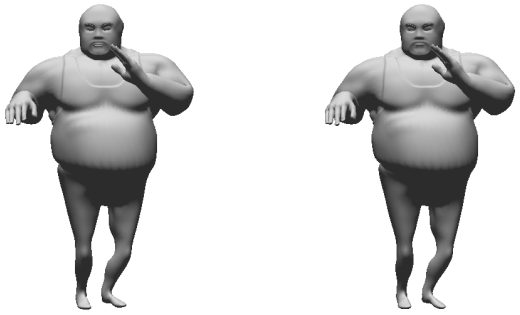

图 4：全身非混合形状绑定的定性结果。原始动画（左）与使用恢复的绑定参数驱动的网格动画（右）之间的差异几乎难以察觉。

该模型使用通过随机组合混合形状生成的网格进行训练，组合时不考虑绑定逻辑。由于绑定逻辑即使相对于范围更有限的因子空间也高度非满射，因此绑定几乎不可能精确表示混合形状的随机组合（图 3）。组合所有混合形状通常会产生不合理的网格，因此每个训练样本中所用混合形状的数量及其各自的取值都遵循均匀分布。每个训练轮次包含 10000 次迭代，每次使用 64 个样本组成的小批次。模型采用与骨骼函数近似相同的优化器和调度策略进行训练。不同的随机权重初始化均得到了相符的结果。

表 2 比较了使用任一种可微绑定近似方法训练绑定逆向近似，并在一个包含 10000 个姿态的测试集上测试所得模型的结果。这些姿态采用与训练集相同的方法随机生成，此外还使用了一组捕获数据。为获得无偏的测试结果，我们不使用绑定函数近似ˆ L_d 进行测试。相反，我们从逆向绑定近似模型中获取每个测试样本的绑定参数向量 r，

表 2：绑定反演结果，显示以毫米为单位的顶点误差。

	平均顶点误差	最大顶点误差
随机姿态		
程序化 $\hat{L}_d$	**2.5**	**15.0**
训练 $\hat{L}_d$	3.0	17.2
[Holden 等人 2015]	9.2	100.0
捕获数据		
程序化 $\hat{L}_d$	**4.8**	**49.0**
训练 L_d	4.9	49.1
[Holden 等人 2015]	10.0	99.0

并将其作为输入传入使用 Maya 动画软件的原始绑定逻辑 L。

使用程序化或训练得到的绑定近似，都比 Holden 先前提出的深度学习方法有显著改进 [Holden 等人 2015]，该方法使用绑定参数真实值进行训练。捕获数据几乎不可能被绑定完美表达，因此无法实现零误差。捕获数据还带有噪声，这解释了所有方法的最大误差都较大。这是有利的，因为逆向骨架技术可以滤除捕获数据中的噪声。

4.3 参数平滑度结果

最后，我们评估该模型对于时间上连贯的输入序列输出平滑绑定参数的实用性。我们使用二阶差分平方作为粗糙度的衡量指标，并将结果与不训练逆向骨架模型、直接优化绑定参数向量的结果进行比较。结果见表 3。我们发现，使用编码器形逆向骨架模型输出的绑定参数平滑度高出一个数量级，而且无需进行噪声增强。尽管如此，在优化之前，每个绑定参数向量都以先前帧的值初始化，以帮助保持时间连贯性。定性来看，直接优化绑定参数向量所得的结果过于不稳定，不适合实际使用。

表 3：展示绑定参数粗糙度的时间连贯性结果

	粗糙度
所提出的方法（程序化） $\hat{L}_d$	3.8e-3
所提出的方法（训练） $\hat{L}_d$	**3.4e-3**
直接使用程序化方法进行优化 $\hat{L}_d$	2.3e-2
直接使用训练方法进行优化 $\hat{L}_d$	2.7e-2

5 结论与未来工作

我们展示了使用可微骨骼函数近似来训练骨骼函数近似的有效性。我们所提出的方法较以往工作有多项改进。首先也是最重要的一点，所提出的方法全面解决了复杂绑定函数的非双射特性，同时还能泛化到无法通过绑定函数精确生成的合理网格。所提出的方法也不需要由艺术家制作的绑定姿势或动画数据集。作为未来工作，我们希望采用针对网格数据而非多层感知机量身设计的架构，以提高准确度并减少训练轮次。

致谢

我们谨此感谢 SEED Electronic Arts 的 Mattias Teye 对本项目作出的宝贵贡献。我们也要感谢 AdobeTM 提供摔跤手角色。

参考文献

Graham Fyffe、Andrew Jones、Oleg Alexander、Ryosuke Ichikari 和 Paul Debevec。2015。利用视频表演捕捉驱动高分辨率面部扫描。*ACM Trans. Graph.* 34，1，文章 8（2015 年 12 月），共 14 页。https://doi.org/10.1145/2638549 Abhijeet Ghosh、Graham Fyffe、Borom Tunwattanapong、Jay Busch、Xueming Yu 和 Paul Debevec。2011。利用偏振球面梯度照明进行多视角人脸捕捉。*ACM Trans. Graph.* 30，6（2011 年 12 月），1–10。https://doi.org/10.1145/ 2070781.2024163 Ian Goodfellow、Yoshua Bengio 和 Aaron Courville。2016。深度学习。MIT出版社，美国马萨诸塞州剑桥，507。http://www.deeplearningbook.org。Stephen Gustafson、Aaron Lo 和 Paul Kanyuk。2020。以解析方式学习逆向绑定映射。载于 *ACM SIGGRAPH 2020*演讲（线上活动，美国）（*SIGGRAPH '20*）。计算机协会，美国纽约州纽约，文章 27，共 2 页。https://doi.org/10.1145/3388767.3407316 Fabian Hahn、Sebastian Martin、Bernhard Thomaszewski、Robert Sumner、Stelian Coros 和 Markus Gross。2012。绑定空间物理。*ACM Transactions on Graphics - TOG* 31（2012 年 7 月）。https://doi.org/10.1145/2185520.2185568 Daniel Holden、Jun Saito 和 Taku Komura。2015。为角色动画学习逆向绑定映射。载于第*14*届*ACM SIGGRAPH / Eurographics*计算机动画研讨会论文集（美国加利福尼亚州洛杉矶）（*SCA '15*）。计算机协会，美国纽约州纽约，165–173。https://doi.org/10.1145/ 2786784.2786788 Daniel Holden、Jun Saito 和 Taku Komura。2017。通过非线性回归学习逆向绑定映射。*IEEE*可视化与计算机图形学汇刊 23，3（2017 年 3 月 1 日），1167–1178。https://doi.org/10.1109/TVCG.2016.2628036 Vinod Nair 和 Geoffrey E. Hinton。2010。整流线性单元改进受限玻尔兹曼机。载于第*27*届国际机器学习会议论文集（以色列海法）（*ICML'10*）。Omnipress，美国威斯康星州麦迪逊，807–814。Stevo Rackovic、Claudia Soares、Dusan Jakovetic 和 Zoranka Desnica。2021。准确、可解释且快速的动画：一种迭代式、稀疏且非凸的方法。https://doi.org/10.48550/ARXIV.2109.08356 Stevo Racković、Cláudia Soares、Dušan Jakovetić、Zoranka Desnica 和 Relja Ljubobratović。2021。混合变形人脸模型的聚类。https://doi.org/10.48550/ ARXIV.2110.15313 Yeongho Seol 和 J. P. Lewis。2014。在动作捕捉流程中调整面部动画。载于*ACM SIGGRAPH 2014*演讲（加拿大温哥华）（*SIGGRAPH '14*）。计算机协会，美国纽约州纽约，文章 13，共 1 页。https://doi.org/10. 1145/2614106.2614108 Steven L. Song、Weiqi Shi 和 Michael Reed。2020。利用深度微分子空间重建实现准确的人脸绑定近似。*ACM Trans. Graph.* 39，4，文章 34（2020 年 7 月），共 12 页。https://doi.org/10.1145/3386569.3392491 Muhammad Usama 和 Dong Eui Chang。2018。迈向具有利普希茨连续性的稳健神经网络。https://doi.org/10.48550/ARXIV.1811.09008